2018-08-13

Frameworks and quality measures used for debriefing in team-based simulation: a systematic review

Endacott, Ruth

http://hdl.handle.net/10026.1/12379

10.1136/bmjstel-2017-000297
BMJ Simulation and Technology Enhanced Learning
BMJ

Frameworks and quality measures used for debriefing in team-based simulation: a systematic review

Accepted for publication in:

BMJ Simulation and Technology Enhanced Learning

DOI: 0.1136/bmjstel-2017-000297

Ruth Endacott[1, 2], RN, PhD, MA, CertEd

Thomas Gale[3,4], BM BS, MEd

Anita O'Connor[1], MSc

Samantha Dix[2], RN, MN, GradCert Health Prof Ed

1. School of Nursing and Midwifery, Plymouth University, Plymouth UK

2. School of Nursing & Midwifery, Monash University, Frankston, Australia

3. Collaboration for the Advancement of Medical Education Research and Assessment, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK

4. Department of Anaesthesia, University Hospitals Plymouth NHS Trust, Plymouth, UK

Corresponding author:

Ruth Endacott, School of Nursing and Midwifery, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK

ruth.endacott@monash.edu

Tel: +44 1752 587488, Fax: +44 1752 586520

Keywords:

Simulation, debriefing, validity, frameworks, interprofessional

Word count:

Abstract	250 words
Main text	4065 words

ABSTRACT

OBJECTIVES

The skill of the debriefer is known to be the strongest independent predictor of the quality of simulation encounters yet educators feel under-prepared for this role. The aim of this review was to identify frameworks used for debriefing team based simulations and measures used to assess debriefing quality.

METHODS

We systematically searched PubMed, CINAHL, MedLine and Embase databases for simulation studies which evaluated a debriefing framework. Two reviewers evaluated study quality and retrieved information regarding study methods, debriefing framework, outcome measures and debriefing quality.

RESULTS A total of 676 papers published between January 2003 and December 2017 were identified using the search protocol. Following screening of abstracts, 37 full-text articles were assessed for eligibility, 26 studies met inclusion criteria for quality appraisal and 18 achieved a sufficiently high quality score for inclusion in the evidence synthesis. A debriefing framework was used in all studies, mostly tailored to the study. Impact of the debrief was measured using satisfaction surveys (n=11) and/or participant performance (n=18). Three themes emerged from the data synthesis: selection and training of facilitators, debrief model and debrief assessment. There was little commonality across studies in terms of participants, experience of faculty, and measures used.

CONCLUSIONS

A range of debriefing frameworks were used in these studies. Some key aspects of debrief for team-based simulation, such as facilitator training, the

inclusion of a reaction phase and the impact of learner characteristics on debrief outcomes, have no or limited evidence and provide opportunities for future research particularly with inter-professional groups.

Background

In simulation learning, debriefing – "a discussion between two or more individuals in which aspects of a performance are explored and analysed with the aim of gaining insights that impact the quality of future clinical practice"[1] - is key and the skill of the debriefer is the strongest independent predictor of overall quality of simulation encounters.[2] In a conceptual paper, Haji et al [3] argued for a distinction between simulation-based and simulation-augmented medical education, with the latter integrating the simulation learning with other educational experiences. This approach also places simulation mainstream, rather than as a special event for the privileged few. Whilst simulation-based education is laudable, simulation is an expensive resource especially when used for small group learning. We therefore need to ensure that learning opportunities are optimised when simulation is used.

Effective interprofessional working is important for standards of patient care and is thought to be highly influenced by the attitudes of health care professionals.[4-6] However, a report from the Centre for the Advancement of Interprofessional Education (CAIPE) highlights that many educators feel underprepared in interprofessional, as compared to uni-professional, settings and recommends that all facilitators receive comprehensive orientation, preparation and on-going support for IPE.[7] Interprofessional team-based simulation allows learning opportunities within the correct educational and professional context[8] and has been shown to improve communication skills and understanding of professional roles.[7] However, debriefing inter-professional groups brings its own unique challenges due to learner differences in background, experience and professional identity[9] requiring faculty to be trained appropriately to debrief interprofessional issues in an effective manner.[8]

Dreifuerst used concept analysis methods to identify defining attributes of debriefing as it relates to simulation,[10] to construct model, borderline and contrary cases and to distinguish between unstructured, structured for critique and structured for reflection approaches to debrief. This is a useful addition to our understanding of debriefing but has yet to be subjected to empirical testing. Previous systematic reviews have focused on the advantages of debrief over no debrief and whether the use of video improves the debrief, [1 11] however there is a lack of research exploring the evidence base underpinning decisions about debriefing. The main aims of this study were to identify (i) frameworks used for debriefing interprofessional and uni-professional team based simulations, (ii) metrics which have been developed to assess the quality of debriefing and (iii) evidence gaps for debrief decisions. The term 'debriefing framework' is used to refer to the structure used for the debriefing discussion.

Methods

Design

A systematic review was conducted following the procedures set out by the Centre for Reviews and Dissemination,[12] whereby specific search terms are used in database searching and papers are selected based on an explicit inclusion and exclusion criteria. We also undertook hand searching of references and sought to identify records through other sources (eg. Google

Scholar), in an attempt to include as many relevant papers as possible in the review. We aimed to identify:

1. debriefing frameworks used for team-based (uni or inter-professional) simulation
2. measures to assess the quality of debriefing

Search strategy

Four electronic databases were searched in December 2017: PubMed, CINAHL, MedLine and Embase. All peer reviewed articles published in English between January 2003 and December 2017 were eligible for inclusion. Our preliminary searches identified many papers that were not relevant. This 15 year window was decided on for pragmatic reasons, and because no relevant papers providing empirical data regarding team-based debriefing were identified prior to this date. As initial searches had identified excessive numbers of papers with either 'framework' or 'method' in the title or abstract, we refined search terms and ran a further search using the keywords: 'Simulation' AND ('Debrief* OR Feedback') AND 'Evaluation' AND ('Quality OR Framework OR Method'). Empirical studies and framework/development studies were included in the review, providing some form of outcome measure was used. Outcome measures assessed quality of the debriefing and/or performance of participants. All included studies used team-based simulation and examined technical and non-technical skills. Studies not published in English, focused on individual debriefing and describing only the quality of the simulation (and not including quality or outcome of the debrief) were excluded.

Quality appraisal

Papers were assessed using the Kmet et al[13] quality appraisal tool. The initial appraisal was conducted by two of the authors, with a third author meeting to discuss any differences in the scoring (RE, TG, AOC, SD). Any discrepancies in scoring were discussed until consensus was reached.

Results

A total of 676 citations were screened; the PRISMA flowchart summarises the review process (Figure 1). Abstracts were reviewed for 253 papers; 41 (6.1%) were found to meet the study criteria after review of titles and abstracts by two authors (RE &AOC or RE & SD). There were no disagreements on inclusion of papers. The remaining 41 full articles were interrogated and assessed for eligibility; 11 were excluded (including concept analysis, application of a theoretical framework and commentary papers).

INSERT FIGURE 1 HERE

A total of 26 papers met the full inclusion criteria and were appraised. Eight papers were excluded from the data synthesis due to a low quality appraisal score (<0.60); this is common in narrative reviews to ensure synthesis of papers of suitable and comparable quality and that recommendations for future practice are not based on low quality evidence. [13] Tables 1 and 2 show the quality appraisal scores for the 26 papers reviewed.

INSERT TABLES 1 AND 2 HERE

A total of 18 papers were included; 1 qualitative study, 15 quantitative studies and 2 studies containing both qualitative and quantitative components. The quantitative KMET scores ranged between 65-100%; the two mixed methods papers[14 15] and the qualitative paper[16] scored 85%. Summary of the 18 included studies is provided at Table 3.

INSERT TABLE 3 HERE

Demographics

There were 2013 participants across the 18 studies (range 9-450). Twelve studies were conducted in the United States, two of which[14 15] contained both qualitative and quantitative components, with the remaining ten comprising quantitative data only. The remaining quantitative studies were conducted in the United Kingdom,[17] Switzerland,[18] Korea[19] and the remaining two in Canada.[20 21] The only wholly qualitative paper included in the review was conducted in the United Kingdom.[16]

Seven studies were conducted with interprofessional teams and four of these examined differences between the professional groups.[16 18 22 23] Geis et al[22] used simulation to model how a new paediatric emergency department would function and to identify latent safety threats; debriefing was structured and included video review. Changes in workload for different professional groups were analysed as the simulated workload of the Department changed. LeFlore et al[23] compared two approaches to Interprofessional team simulation and debriefing; changes in knowledge test scores and satisfaction with the simulation/debrief were reviewed by professional group. In the Freeth et al qualitative study[16] some excerpts from interviews identified participants by professional group but there was no comparison between groups. Kolbe et al[18] found that evaluation of their debriefing model – TeamGAINS – did not differ by job role (nurse or doctor).

Debriefing frameworks

All studies included a structured debriefing framework, mostly tailored to the individual study (see Table 4). Five authors used a previously validated framework: the Ottawa Global Rating Scale,[20] TeamGAINS,[18] Debriefing for Meaningful Learning,[24] Structured and Supported Debriefing[19] and Guided Team Self Correction (GTSC).[25] In 11 studies, outcome measures were used to assess debrief quality (faculty behaviours)[14 15 17 18 22-24 26-29] and in 12 studies change in performance following the debrief was measured (participant behaviours). [16 18 20-25 30-32]

INSERT TABLE 4 HERE

Performance measures

The majority of studies (12/18) used some measure of performance to judge the success of the debriefing framework, using a before and after design or comparing two debriefing frameworks (Table 4). A total of 17 measures were used in the 12 studies (Table 4).

Synthesis

All papers were read in full by two authors; a combination of inductive and deductive thematic analysis was used to develop codes and categories to relevant extracts and organise these findings under main thematic headings.

These are presented at Figure 2. Deductive codes were derived from the review aims and the inductive component allowed codes to emerge from the data. A synthesis of these findings was used to identify key themes.

Several key themes were identified through this synthesis of the findings; two authors discussed these themes until a consensus was reached. These themes were: selection and training of debrief facilitators, debrief model and assessment of debrief. The themes are discussed below; summary of the evidence, and evidence gaps, for each theme is presented at Figure 2.

INSERT FIGURE 2 HERE

Selection and training of debrief facilitators

Most of the studies were conducted with a trained debrief facilitator[15-18 22 24 26 29 31 32] with one research team reporting use of 'PowerPoint plus audio' with no indication whether the 'audio' was pre-recorded or provided by a facilitator. [14] A RCT compared two approaches to debrief: within-team debrief, with a leader from within the team providing the debrief, and instructor led debrief. [20] Team performance, assessed using the Team Emergency Assessment Measure (TEAM), [33] improved following debrief in both groups ($F_{1,38} = 7.93$, p=0.008); there was no significant difference between within-team or instructor debrief ($F_{1,38} = 0.43$, NS p=0.52). Oikawa et al found that self-debriefing was as effective as faculty debriefing in improving self and team performance assessment across 4 sequential scenarios. [32]

Different study designs make it impossible to state that one type of facilitator is superior; performance in individual studies improved when the team leader,[20] instructor,[15] faculty[32] or team member[32] led the debrief. Similarly no studies provided evidence that training actually makes any difference.

Debrief model

The format of debriefing reported in the studies varied in three areas: degree of structure, use of video clips and timing of the debrief.

All authors described a debrief framework, with variation in the detail provided. Three authors specify an initial reaction stage ('how was that for you?'), followed by attention to technical and/or non-technical skills and how they were performed in the simulation scenarios; Lammers et al [15] and Van Heukelom et al [27] refer to this first stage as 'decompression', whilst Kolbe et al [18] describe it as 'reactions'. No one structure was used across studies; most authors tailored an existing debrief framework.

Training faculty to use GTSC to structure the debrief had a significant impact on overall team performance, over traditional debrief methods ($t(11) = 1.98$, $p=<0.05$ (one-tailed)). [25] The group receiving GTSC also developed mental models more similar to those developed by an expert group. In a pre-test post-test study Pediatric Emergency Medicine fellows were trained to use a cardiac arrest debriefing model (REFLECT) with teams of four. The fellows and team members reported significant improvement in use of REFLECT components (63 vs 82%) but blinded expert reviewers reported a non-significant improvement (60 vs 76%). [29]

Use of Cognitive Disposition to Respond (CDR) to structure the debrief, with technical/knowledge based debrief as the control, resulted in higher satisfaction scores for the technical/knowledge based debrief. This did not reach significance.[14] Leflore and Anderson[23] compared a facilitated debrief (Group A) with a modified debrief (Group B) in which time for questions was allowed. However the learning interaction was also different with Group A using self-directed learning and Group B observing experts completing the scenario. Group B had higher satisfaction scores but there is no indication whether this was due to the expert modelling or the modified debrief.

Video clips were included in the debrief in 7 of the studies[15 16 20-23 26] but extent of video use described by the authors was variable. In one study, the researchers compared no debrief (control) with oral debrief (intervention 1) and oral plus video debrief (intervention 2) using a pre-post design with anaesthesia residents.[21] There was significant improvement in total Anaesthesia Non-Technical Skills (ANTS) score ($F_{2,39}$ = 6.10, p=<0.005) and scores in each of the 4 domains for both intervention groups but no significant difference between oral and oral + video groups on total or individual domain scores. Similarly a pre-test post-test study comparing video-assisted debrief with oral debrief alone with nursing students reported a higher mean score on behaviour for those in the video-assisted debrief group than the control group (6.62 vs 4.23), but this did not reach significance.[30]

In most studies, debriefing was conducted at the end of the simulation exercise; the one exception was the study conducted by Van Heukelom et al,[27] who compared in-simulation debrief (identifying learning points and errors as they arise during the simulation) and post-simulation debrief. They report that self-reported confidence and knowledge improved for both groups (Spearman's R = 0.5 with $p \leq 0.001$ for all results) with no significant difference between groups. However, the post-simulation debrief group had significantly higher scores for three items on the debriefing satisfaction scale. In seven studies, participants completed a further simulation scenario following the debrief;[20-25 30] this is reviewed in detail below.

The studies reviewed provide evidence that debriefing frameworks can improve outcomes; however there is no evidence that including a reaction phase or using video makes any difference to outcomes.

Assessment of the debrief

There were two approaches to assessment of debrief: assessment of debrief quality and change in performance following the debrief.

The quality of the debrief was assessed through satisfaction scores or through analysis of debrief videos. Satisfaction was rated by participants[14 23 24 27 28] or faculty,[26] or both.[17 18 29] Kolbe et al[18] also measured psychological safety and leader inclusiveness before and after the debrief and found both measures significantly improved ($t(59)=-2.26$, $p=0.028$ and $t(60)=-2.07$, $p=0.048$). In four studies, analysis of debrief videos was conducted using an existing tool: Brett-Fleegler et al[26] used the Debriefing Assessment for Simulation in Healthcare (DASH) with 114 simulation instructors to test validity and reliability and

Lammers et al[15] used a Root Cause Analysis (RCA) framework to examine the quality of RCA processes in a simulated pre-hospital paediatric emergency. Hull et al[17] used Objective Structured Assessment of Debriefing (OSAD) with expert debriefing evaluators and faculty debriefing and Zinns et al[29] used the REFLECT post resuscitation debriefing framework.

Significant improvement in performance following debrief was reported in several studies. Change in performance was assessed using (i) a (different) simulation scenario conducted after the debrief,[20=23 25] (ii) participant knowledge, assessed using a pre/post knowledge test, [25] (iii) participant self-reported confidence and knowledge,[27] and (iii) mental model accuracy.[25]

The post-debrief simulation performance was assessed using a range of existing measures: the Mayo High Performing Team Scale,[22] the Team Emergency Assessment Measure (TEAM),[20] Anaesthesia Non-Technical Skills (ANTS),[21] Behaviour Assessment Tool, based on CRM principles and validated in previous studies by the authors,[23] the Health Sciences Reasoning Test,[24] Team Dynamics[31] and Team Clinical Performance.[31] In the Geis et al study,[22] the Phase 1 (pre-debriefing) simulation was conducted in the simulation lab and the Phase 2 (post-debriefing) was conducted in the hospital, hence change in behaviour could not be attributed solely to the debrief.

Despite some studies using more than one performance measure, none of the studies reported correlations across performance measures. Where performance data was analysed in the context of demographic data items, these were mainly limited to professional group[16 18 22 23] and work experience.

Discussion

There was little commonality across the papers in terms of participants, experience of faculty and measures used, however all studies used a debriefing framework to provide structure for the debriefs often underpinned by theoretically derived methods to facilitate interaction of participants. 18 different debriefing frameworks were described, showing divergence in preferred debriefing techniques and strategies amongst the studies but the frameworks commonly started with a “reaction” or “decompression” phase to encourage self / team reflection. The reaction phase assumes that participants will “let off steam” during the first few minutes of a simulation debrief which provides facilitators with content which should be discussed at some stage in the debrief but also allows participants to express their emotions straight away and provide a more balanced environment for objective reflection later in the debrief.[18] None of the studies compared this reaction phase with no reaction phase so the impact is unknown. All debriefing frameworks covered either technical or non-technical aspects, or both and some studies compared participant reactions to either technical / non-technical aspects. Non-technical skills were addressed through the use of expert models such as crisis resource management principles or through techniques such as CDR and Advocacy Inquiry (AI) aimed at identifying mental models of participants which lead to certain behaviours.[14 26] Bond[14] found that technical debriefing was better received by participants than cognitive debriefing, although Dreifuerst[34] reported that learners prefer debrief with reflection.

The debriefing model described by Kolbe and colleagues[18] reflects the recommendations of several earlier authors and comprises 6 steps: reactions; debrief clinical component; transfer from simulation to reality; reintroduce the expert model; summarise the debriefing, and practice/improve clinical skills as required. This model, as a whole, was shown to have some benefits but our review has shown varying degrees of evidence for each of these steps, as illustrated at Figure 2.

Debriefing theory

Different techniques are used to focus the debrief on individuals and team members as well as observers. Debriefing models utilised a range of theoretical techniques to facilitate interaction of the whole group through guided team self-correction, peer assessment, self and team reflection. [18 23 25 30-32] Guided team self-correction and circular questioning[18 25] are techniques which switch the focus to the whole team and encourage active participation and reflexivity from all members of the group. Smith-Jentsch et al developed the technique of guided team self-correction (GTSC) where members of the team are responsible for identifying their own team performance problems plus process orientated goals for improvement.[25] In GTSC, an expert model of teamwork is used as an organisational framework at the briefing and then debriefing stages when participants are asked to discuss both positive and negative examples of each component. Debriefing theory developed by Salas and colleagues makes the assumption that the use of an expert model provides a common language for participants to use during team debriefs which helps to form shared team mental models which match the expert framework.[25 35] Reflecting on both positive and negative examples of behaviour has been found to develop stronger mental models and focusing on a few critical performance issues to identify learner 'process orientated goals' helps to ensure that learning is not scenario specific. High level facilitation allows participants to contribute to the majority of discussion in the debrief which maximises individual reflection and team based reflexivity so that the learners are reaching a deeper level of understanding about the interactions which have taken place, rather than listening to expert opinion by the debriefer. With techniques such as GTSC, the debriefer facilitates from a non-judgmental perspective without expressing their own expert opinion until the latter stages of the debrief, if at all.

In contrast, Advocacy Inquiry (AI) is more instructor led where the debriefer will highlight a performance gap encountered by an individual during the simulation and use direct questioning to uncover underlying mental frames which led to certain actions or behaviours.[18, 26] The conceptual framework and underlying theory assumes that by exploring the mental frames or thought processes which have led to certain behaviours, the learner is able to rewire these thought processes for similar situations in the future, resulting in different actions or interactions.[36]

A central tenet across debriefing theories for teams is the development of a shared understanding across participants and facilitator. However, the seven studies we reviewed that were conducted with interprofessional teams did not appear to test mental model consistency across professions.

Learning environment

Creating the right environment has been eloquently described as a 'task-relationship dilemma' [36 37] between the need to provide honest feedback on the task without damaging the relationship between teacher and learner. The studies included in our review suggest that greater attention is being paid to this, as evidenced by validation of measures for the assessment of perceived psychological safety[18] and in the debriefing and evaluation of satisfaction.[14 23 26 27] The use of video as part of the debrief is not supported by studies included in our review; this is consistent with an earlier meta-analysis. [1]

Training of debriefers

The majority of studies used trained debrief facilitators to conduct the debrief although two studies showed that self-debrief within teams was as effective as instructor led debrief.[20 32] Cheng and colleagues, in their systematic review of debriefing features, outcomes and effectiveness,[1] found that there may be benefits in expert modelling, although meta-analysis of relevant studies revealed non-significant effects.

When instructors perform debriefs, in-simulation debriefing does not work as well as post simulation debriefing.[27] A study examining student perceptions of debriefing[38] also revealed that students prefer debriefing immediately following the simulation and that timing was more important than the debriefing model. However, comparison of studies by Cheng and colleagues[1] suggest that factors such as task complexity and individual or team-based learning may be better indicators for the timing of debriefing. Further training in specific techniques such as GTSC and CDR raises the quality of debriefings, so it is important to use experienced facilitators, an agreed / previously validated debriefing framework and to supplement facilitator training with technique-specific instruction to optimise debriefing quality. Standards of best practice for simulation[39] advocate that the debrief facilitator has specific training and has witnessed the simulation activity. Debriefing frameworks encourage facilitators to focus on a few critical issues, include a range of formats and address technical and cognitive aspects, non-technical skills, and transfer of learning into practice.

Quality metrics

We identified 4 previously validated metrics used to measure the quality of debriefs; DASH, OSAD, Reflect and DES, with DASH and OSAD the preferred metric in more than one study. These metrics utilise faculty, participant or objective raters to score aspects of faculty performance except the DES which assesses participant feelings as a result of the debriefing experience. Whilst these instruments have good evidence of reliability and validity, further studies are needed to establish validity in different contexts and compare the utility of different tools.

Integration with previous work

Previous systematic reviews have shed light on the advantages of debrief over no debrief and the lack of evidence that the use of video improves the debrief.[1 11] Our review supports both of these findings. Methods of debriefing have been reviewed in previous narrative reviews [2 38] and systematic reviews.[1 11] Of note Cheng and colleagues were only able to conduct meta-analysis on a small number of the 177 studies included in their systematic review, due to incomplete reporting by researchers.[1] In a more theoretical approach, the defining attributes of debriefing identified by Dreifuerst[10] - reflection, emotion, reception, and integration and assimilation[10] – enabled the author to identify model, borderline and contrary cases, in line with the concept analysis method.[40]

The main contribution of this systematic review has been to identify debriefing frameworks some of which have been validated in various contexts using theoretical approaches. However the number of bespoke frameworks used highlights the diversity of debriefing practice and approaches to outcome measurement, and that more work should be done to compare debriefing frameworks in order to develop evidence for best practice.

Implications for current practice and future research

Our review suggests that the use of a debrief framework improves debrief quality, subsequent behaviours and teamwork performance. The findings strongly support the use of a validated debrief framework by debriefers but investment in preparation of the faculty is also important, to supplement facilitator training with technique-specific instruction to optimise debriefing quality. Further research is needed to validate measures of debrief quality in different contexts and outcome measures following debriefing. The number of bespoke instruments used across the studies illustrates the difficulty with conducting reviews such as this, particularly with limitations to meta-analysis. It would be worth considering whether there are key outcomes (and associated outcome measures) that should be considered good practice for simulation research, similar to the core outcomes dataset approach being promulgated for clinical research (http://www.comet-initiative.org/).

Some key aspects of debrief for team-based simulation, such as facilitator training, the inclusion of a reaction phase and the impact of learner characteristics on debrief outcomes, have no or limited evidence and provide opportunities for future research, particularly with interprofessional groups.

Funding statement: This work was supported by the UK Higher Education Authority Teaching Development Grant number GEN-620.

Contributors: RE and TG contributed substantially to conception and design of the study. AOC conducted the searches; RE, TG, AOC and SD conducted critical appraisal of the papers. RE, TG and SD interpreted the review papers and drafted the manuscript. All authors have had access to the review findings and take responsibility for accuracy and integrity of the work. All authors critically appraised the content and approved the final version.

Competing interests: None declared.

References

1. Cheng A, Eppich W, Grant V, Sherbino J, Zendejas B, Cook DA. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Medical Education*. 2014; **48**(7):657-66.

2. Fanning RM, Gaba DM. The role of debriefing in simulation based learning. Simulation in Healthcare 2007;**2**(2):115-25.

3. Haji, F. A., Hoppe, D. J., Morin, M. P., Giannoulakis, K., Koh, J., Rojas, D., & Cheung, J. J. H. (2013). What we call what we do affects how we do it: a new nomenclature for simulation research in medical education. *Advances in Health Sciences Education*, 1–8. doi:10.1007/s10459-013-9452-x.
4. Baker DP, Gustafson S, Beaubien J, Salas E, Barach P. (2005) Medical teamwork and patient safety: the evidence-based relation. *AHRQ publication*. 5(53):1-64.
5. Hind M., Norman I., Cooper S., Gill E., Hilton R., Judd P. & Jones SC (2003). Interprofessional perceptions of health care students. *Journal of Interprofessional Care* 17(1), 21-34
6. Thistlethwaite J, Moran M. Learning outcomes for interprofessional education (IPE): Literature review and synthesis. *Journal of Interprofessional Care*. 2010;**24** (5):503-13.
7. Barr H, Low H. Interprofessional education in preregistration courses: A CAIPE Guide for Commissioners and Regulators of Education. CAIPE, Fareham 2012. http://caipe.org.uk/silo/files/caipe-guide-for-commissioners-nd-regulators-of-eduction-.pdf

8. Boet S, Bould MD, Layat Burn C, Reeves S. Twelve tips for a successful interprofessional team-based high-fidelity simulation education session. *Medical Teacher* 2014;**36**(10):853-57.

9. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation based medical education research: 2003-2009. Medical Education 2010;**44**(1):50-63.

10. Dreifuerst KT. The essentials of debriefing in simulation learning: a concept analysis. *Nurs Educ Perspect* 2009;**30**:109-14

11. Levett-Jones T, Lapkin S. A systematic review of the effectiveness of simulation debriefing in health professional education. *Nurse Education Today*. 2014; **34**(6):e58-63.

12. Centre for Reviews and Dissemination (CRD). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination; 2009.

13. Kmet, LM, Lee RC, Cook LS. *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields*. Alberta Heritage Foundation for Medical Research. Canada, 2004.

14. Bond WF, Deitrick LM, Eberhardt M, Barr GC, Kane BG, Worrilow CC, Arnold DC, Croskerry P. Cognitive versus technical debriefing after simulation training. *Academic Emergency Medicine*. 2006;**13**(3):276-83.

15. Lammers R, Byrwa M, Fales W. Root causes of errors in a simulated prehospital pediatric emergency. *Academic Emergency Medicine*. 2012;**19**(1):37-47.

16. Freeth D, Ayida G, Berridge EJ, Mackintosh N, Norris B, Sadler C, Strachan A. Multidisciplinary obstetric simulated emergency scenarios (MOSES): Promoting patient safety in obstetrics with teamwork-focused interprofessional simulations. *Journal of Continuing Education in the Health Professions*. 2009; **29**(2):98-104

17. Hull L, Russ S, Ahmed M, Sevdalis N, Birnbach DJ. Quality of interdisciplinary postsimulation debriefing: 360° evaluation. *BMJ Simulation and Technology Enhanced Learning.* 2017; 3(1):9-16.

18. Kolbe M, Weiss M, Grote G, Knauth A, Dambach M, Spahn DR, Grande B. TeamGAINS: a tool for structured debriefings for simulation-based team trainings. *BMJ Quality & Safety* 2013; **22**: 541-553.

19. Kim JH, Kim YM, Park SH, Ju EA, Choi SM, Hong TY. Focused and Corrective Feedback Versus Structured and Supported Debriefing in a Simulation-Based Cardiac Arrest Team Training: A Pilot Randomized Controlled Study. *Simul Healthc.* 2017; **12**(3):157-164.

20. Boet S, Bould MD, Sharma B, Revees S, Naik VN, Triby E, Grantcharov T. Within-team debriefing versus instructor-led debriefing for simulation-based education: a randomized controlled trial. *Annals of Surgery* 2013; **258**(1):53-8.

21. Savoldelli GL, Naik VN, Park J, Joo HS, Chow R, Hamstra SJ. Value of Debriefing during Simulated Crisis Management Oral versus Video-assisted Oral Feedback. *Anesthesiology* 2006;**105**(2):279-85.

22. Geis GL, Pio B, Pendergrass TL, Moyer MR, Patterson MD. Simulation to assess the safety of new healthcare teams and new facilities. *Simulation in Healthcare*. 2011; **6**(3):125-33.

23. LeFlore JL, Anderson M. Alternative educational models for interdisciplinary student teams. *Simulation in Healthcare*. 2009; **4**(3):135-42.

24. Forneris SG, Neal DO, Tiffany J, Kuehn MB, Meyer HM, Blazovich LM, Holland AE, Smerillo M. Enhancing clinical reasoning through simulation debriefing: A multisite study. *Nursing education perspectives*. 2015; 36(5):304-10.

25. Smith-Jentsch KA, Cannon-Bowers JA, Tannenbaum SI, Salas E. Guided team self-correction impacts on team mental models, processes, and effectiveness. *Small Group Research*. 2008; **39**(3):303-27.

26. Brett-Fleegler M, Rudolph J, Eppich W, Monuteaux M, Fleegler E, Cheng A, Simon R. Debriefing assessment for simulation in healthcare: development and psychometric properties. *Simulation in Healthcare* 2012; **7**:288-94.

27. Van Heukelom JN, Begaz T, Treat R. Comparison of postsimulation debriefing versus in-simulation debriefing in medical simulation. *Simulation in Healthcare*. 2010; **5**(2):91-7.

28. Reed SJ. Written debriefing: Evaluating the impact of the addition of a written component when debriefing simulations. *Nurse Education in Practice*. 2015; 15(6):543-8

29. Zinns LE, Mullan PC, O'Connell KJ, Ryan LM, Wratney AT. An Evaluation of a New Debriefing Framework: REFLECT. *Pediatric Emergency Care*. 2017 Jun 14.

30. Grant JS, Dawkins D, Molhook L, Keltner NL, Vance DE. Comparing the effectiveness of video-assisted oral debriefing and oral debriefing alone on behaviors by undergraduate nursing students during high-fidelity simulation. *Nurse Education in Practice*. 2014; 14(5):479-84.

31. Kim JH, Kim YM, Park SH, Ju EA, Choi SM, Hong TY. Focused and Corrective Feedback Versus Structured and Supported Debriefing in a Simulation-Based Cardiac Arrest Team Training: A Pilot Randomized Controlled Study. *Simul Healthc*. 2017; 12(3):157-164.

32. Oikawa S, Berg B, Turban J, Vincent D, Mandai Y, Birkmire-Peters D. Self-Debriefing vs Instructor Debriefing in a Pre-Internship Simulation Curriculum: Night on Call. *Hawaii J Med Public Health*. 2016; 75(5):127-32.

33. Cooper S, Cant R, Porter J et al. Rating medical emergency teamwork performance: development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010; **81**: 446-452.

34. Dreifuerst KT. Using debreifing for meaningful learning to foster development of clinical reasoning in simulation. *J Nurs Educ* 2012;**51**:326-33.

35. Salas E, Klein C, King H, Salisbury M, Augenstein JS, Birnbach DJ, Robinson DW, Upshaw C. (2008) Debriefing medical teams: 12 evidence-based best practices and tips. Joint Commission Journal on Quality and Patient Safety. 34(9):518-27.

36. Rudolph JW, Simon R, Rivard P, Dufresne RL, Raemer DB. Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Aesthesiology Clin* 2007; **25**: 361-376.

37. Rudolph JW, Foldy EG, Robinson T, Kendall S, Taylor SS, Simon R. Helping without harming. The instructor's feedback dilemma in debriefing – a case study. *Simul Healthcare* 2013; **8**: 304-316.

38. Cantrell MA. The importance of debriefing in clinical simulations. *Clinical Simulation in Nursing* 2008; **4**(2): e19-e23.

39. The INACSL Board of Directors. Standard VI: the debriefing process. *Clinical Simulation in Nursing* 2011; **7**, s16-s17.

40. Walker LO, Avant KC. *Strategies for theory construction in nursing* (4th ed.). Upper Saddle River, NJ: Prentice Hall, 2005.

41. Auerbach M, Kessler D, Foltin JC. Repetitive pediatric simulation resuscitation training. *Pediatric Emergency Care*. 2011; **27,** 29-31.

42. Cheng A, Goldman RD, Aish MA, Kissoon N. A simulation-based acute care curriculum for pediatric emergency medicine fellowship training programs. *Pediatric Emergency Care*. 2010; **26**,475-80.

43. Cooper JB, Singer SJ, Hayes J, Sales M, Vogt JW, Raemer D, Meyer GS. Design and evaluation of simulation scenarios for a program introducing

patient safety, teamwork, safety leadership, and simulation to healthcare leaders and managers. *Simulation in Healthcare.* 2011; **6**, 231-8.

44. Kable AK, Arthur C, Levett-Jones T, Reid-Searl K. Student evaluation of simulation in undergraduate nursing programs in Australia using quality indicators. *Nursing & Health Sciences.* 2013; **15**, 235-43.

45. Kuiper R, Heinrich C, Matthias A, Graham MJ, Bell-Kotwall L. Debriefing with the OPT model of clinical reasoning during high fidelity patient simulation. *International Journal of Nursing Education Scholarship.* 2008; **5**, 1-4.

46. Morrison AM, Catanzaro AM. High-Fidelity Simulation and Emergency Preparedness. *Public Health Nursing.* 2010; **27**, 164-73.

47. West E, Holmes J, Zidek C, Edwards T. Intraprofessional collaboration through an unfolding case and the just culture model. *Journal of Nursing Education*. 2013; **52**, 470-4.

48. Wetzel EA, Lang TR, Pendergrass TL, Taylor RG, Geis GL. Identification of latent safety threats using high-fidelity simulation-based training with multidisciplinary neonatology teams. Joint Commission journal on quality and patient safety. 2013; **39**, AP1-3.

Table 1 Quality appraisal of quantitative studies

Papers	Auerback et al (2011)[41]	Boet et al. (2013)[20]	Bond et al. (2006)[14]	Brett-Fleegler et al. (2012)[26]	Cheng et al. (2010)[42]	Cooper et al. (2011)[43]	Forneris et al (2015)[24]	Geis et al. (2011)[22]	Grant et al (2014)[30]
Question / objective sufficiently described?	2	2	2	2	1	1	1	2	1
Study design evident and appropriate?	2	2	2	2	1	2	2	2	2
Method of subject / comparison group selection or source of information / input variables described and appropriate?	1	2	1	N/A	1	2	1	1	1
Subject (and comparison group) characteristics sufficiently described?	1	2	0	N/A	0	1	1	1	1
If interventional and random allocation was possible, was it described?	0	2	2	N/A	N/A	2	1	N/A	1
If interventional and blinding of investigators was possible, was it reported?	0	2	2	N/A	N/A	N/A	N/A	N/A	2
If interventional and blinding of subjects was possible, was it reported?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Outcome and exposure measure(s) well defined and robust to measurement / misclassification bias?	1	2	2	1	2	1	2	2	2
Sample size appropriate?	1	2	1	1	0	1	2	1	1
Analytic methods described / justified and appropriate?	1	2	1	2	0	1	2	2	2
Some estimate of variance is reported for the main results?	1	2	2	2	0	0	2	2	2
Controlled for confounding?	0	2	0	1	0	1	0	1	0
Results reported in sufficient detail?	1	2	1	2	1	1	2	2	2
Conclusions supported by the results?	1	2	1	2	1	1	1	2	2
Summary score	**0.46**	**1.00**	**0.65**	**0.83**	**0.32**	**0.58**	**0.71**	**0.82**	**0.73**

Table 1 (contd) Quality appraisal scores for Quantitative Studies

Papers	Hull et al (2017)[17]	Kable et al. (2013)[44]	Kim et al (2017)[19]	Kolbe et al. (2013)[18]	Kuiper et al. (2008)[45]	Lammers et al. (2012)[15]	LeFlore & Anderson (2009)[23]	Morrison & Catanzaro (2010)[46]
Question / objective sufficiently described?	2	1	2	2	1	2	2	1
Study design evident and appropriate?	2	1	2	2	1	2	2	1
Method of subject / comparison group selection or source of information / input variables described and appropriate?	1	1	1	1	1	2	1	1
Subject (and comparison group, if applicable) characteristics sufficiently described?	0	0	2	2	1	2	2	0
If interventional and random allocation was possible, was it described?	N/A	N/A	2	N/A	N/A	N/A	1	1
If interventional and blinding of investigators was possible, was it reported?	N/A	N/A	2	N/A	N/A	N/A	2	N/A
If interventional and blinding of subjects was possible, was it reported?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias?	2	1	2	2	1	2	2	1
Sample size appropriate?	2	1	1	1	1	2	1	1
Analytic methods described / justified and appropriate?	2	2	2	1	2	2	2	1
Some estimate of variance is reported for the main results?	1	2	2	2	0	2	2	N/A
Controlled for confounding?	N/A	0	2	1	1	1	1	N/A
Results reported in sufficient detail?	2	2	2	2	1	1	2	1
Conclusions supported by the results?	2	1	2	2	1	2	2	1
Summary score	**0.80**	**0.55**	**0.92**	**0.82**	**0.50**	**0.91**	**0.85**	**0.45**

Table 1 (contd) Quality appraisal scores for Quantitative Studies

Papers	Oikawa et al (2016)[32]	Reed (2015)[28]	Salvodelli et al (2006)[21]	Smith-Jentsch et al. (2008)[25]	Van Heukelom et al. (2010)[27]	West et al. (2013)[47]	Wetzel et al. (2013)[48]	Zinns et al (2017)[28]
Question / objective sufficiently described?	2	1	2	2	2	0	1	1
Study design evident and appropriate?	2	2	2	2	2	1	1	2
Method of subject / comparison group selection or source of information / input variables described and appropriate?	1	1	2	2	2	0	1	1
Subject (and comparison group, if applicable) characteristics sufficiently described?	0	0	2	1	1	0	1	0
If interventional and random allocation was possible, was it described?	1	2	2	1	2	0	N/A	N/A
If interventional and blinding of investigators was possible, was it reported?	N/A	N/A	2	1	0	N/A	N/A	2
If interventional and blinding of subjects was possible, was it reported?	2	2	0	0	N/A	N/A	N/A	N/A
Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias?	2	2	2	2	2	1	1	2
Sample size appropriate?	1	1	2	1	2	1	1	1
Analytic methods described / justified and appropriate?	2	2	2	2	2	1	1	2
Some estimate of variance is reported for the main results?	2	2	2	2	2	N/A	0	2
Controlled for confounding?	1	1	2	1	1	N/A	N/A	N/A
Results reported in sufficient detail?	2	2	2	2	2	0	1	1
Conclusions supported by the results?	2	2	2	2	1	1	1	1
Summary score	**0.77**	**0.77**	**0.93**	**0.75**	**0.81**	**0.25**	**0.45**	**0.68**

Table 2 Quality appraisal scores for Qualitative Studies

Papers	Bond et al. (2006)[14]	Freeth et al. (2009)[16]	Lammers et al. (2012)[15]
Question / objective sufficiently described?	2	2	2
Study design evident and appropriate?	2	2	2
Context for the study clear?	2	2	2
Connection to a theoretical framework / wider body of knowledge?	2	2	1
Sampling strategy described, relevant and justified?	1	1	1
Data collection methods clearly described and systematic?	2	1	2
Data analysis clearly described and systematic?	2	2	1
Use of verification procedure(s) to establish credibility?	2	2	2
Conclusions supported by the results?	1	2	2
Reflexivity of the account?	1	1	2
Summary score	**0.85**	**0.85**	**0.85**

Table 3 Summary of studies included in the narrative synthesis

	Reference/country	*Aim*	*Study Design*	*Participants and sample*	*Findings*
1	Boet et al. (2013) Canada	Compare effectiveness of an interprofessional within-team debriefing with instructor-led debriefing on team performance during simulated crisis.	Randomized, controlled, repeated-measures design. Teams randomized to within-team or instructor-led debriefing groups. After debriefing, teams managed different post-test crisis scenario. Sessions were video -taped and blinded expert examiners used TEAM scale to assess performance.	N=120 (40 teams made up of 1 anaesthesia trainee, 1 surgical trainee, 1 staff circulating operating room nurse).	Team performance significantly improved from pre- to post-test, regardless of type of debriefing ($F_{1,38}$ = 7.93, p=0.008). No significant difference in improvement between within-team or instructor-led debriefing.
2	Bond et al. (2006) United States	To assess learner perception of high-fidelity mannequin-based simulation and debriefing to improve understanding of 'cognitive dispositions to respond' (CDRs).	EM residents exposed to two simulations and block-randomized to technical/knowledge debriefing before completing written survey and interview with ethnographer. Four investigators reviewed interview transcripts and qualitatively analysed comments.	N=62 emergency medicine (EM) residents.	Technical debriefing was better received than cognitive debriefing. Authors theorize that an understanding of CDRs can be facilitated through simulation training.
3	Brett-Fleegler et al. (2012) United States	Examine reliability of DASH (Debriefing Assessment for Simulation in Healthcare) scores in evaluating quality of health care simulation debriefings and whether scores demonstrate evidence of	Rater trainees familiarised with DASH before watching, rating and then discussing 3 separate course introductions and subsequent debriefings. Inter-rater reliability, intraclass correlations and internal consistency were calculated.	N=114 international health care educators participated in 4.5 hour Web-based interactive DASH rater training sessions (nurses, physicians, other health professionals and Masters and PhD	Differences between the ratings of the 3 standardized debriefings were statistically significant p<0.001. DASH scores showed evidence of good reliability and preliminary evidence of validity.

		validity.		educators).	
4	Forneris et al (2015) United States	To investigate the impact of Debriefing for Meaningful Learning (DML) on clinical reasoning	Quasi-experimental pretest-posttest repeated measure design. Teams randomly assigned to DML or usual debriefing. Clinical reasoning was evaluated using the Health Sciences Reasoning Test (HSRT).	N=153 UG nursing students (teams of 4)	Significant improvement in HSRT mean scores for the intervention group (p=0.03) with control group NS. The change in HSRT mean scores between the intervention and control groups was not significant (p=0.09).
5	Freeth et al. (2009) United Kingdom	Examination of participants perceptions of MOSES (the multidisciplinary obstetric simulated emergency scenarios course) designed to enhance NTS among obstetric teams/improve patient safety.	Telephone (47) or email (8) interviews with MOSES course participants and facilitators and analysis of video-recorded debriefings.	N=93 (Senior midwives n=57, obstetricians n=21, obstetric anesthetists n=15).	Many participants improved their knowledge and understanding of interprofessional team working, especially communication and leadership in obstetric crisis situations. Participants with some insight into their non-technical skills showed the greatest benefit in learning. Interprofessional simulation is a valuable approach to enhancing nontechnical skills.
6	Geis et al. (2011) United States	Define optimal health care team roles and responsibilities, identify latent safety threats (LSTs) within the new	Prospective pilot investigation using laboratory and in situ simulations totalling 24 critical patient scenarios conducted over four sessions (over 3 months).	N=81 Health care providers (predominantly nurses, paramedics and physicians).	MHPTS means were calculated for each phase of training. Simulation laboratory teamwork scores showed a mean of 18.1 for the first

		environment and screen for unintended consequences of proposed solutions.			session and 18.9 for the second session (*P*=0.68). In situ teamwork scores showed a mean of 12.3 for the first session and 15 for the second session (*P* = 0.25). Overall laboratory mean was 18.5 (SD 2.31) compared with overall in situ mean of 13.7 (SD 4.40), indicating worse teamwork during in situ simulation (*P* = 0.008).
7	Grant et al (2014) United States	To compare the effectiveness of video-assisted oral debriefing (VAOD) and oral debriefing alone (ODA) on participant behaviour	Quasi-experimental pre-test, post-test design. Teams were randomised to intervention (VAOD) or control (ODA). Behaviours were assessed using adapted Clinical Simulation Tool (CSET).	N= 48 UG nursing students: 24 intervention and 24 control (teams of 4 or 5 students)	The VAOD group had higher mean score (6.62, SD 6.07) than the control group (4.23, SD 4.02) but this did not reach significance (*p*=0.11).
8	Hull et al (2017) UK	To explore the value of 360° evaluation of debriefing by examining expert debriefing evaluators, debriefers and learners' perceptions of the quality of interdisciplinary debriefings.	Cross-sectional observational study. The quality of debriefing was assessed using the validated Objective Structured Assessment of Debriefing (OSAD) framework.	N= 278 students, in 41 teams	Expert debriefing evaluators and debriefers' perceptions of debriefing quality differed significantly; debriefers perceived the quality of debriefing they provided more favourably than expert debriefing evaluators. Learner perceptions of the quality of debriefing differed from both expert evaluators and

					debriefers' perceptions.
9	Kim et al (2017) Korea	To compare the educational impact of two postsimulation debriefing methods: (focused and corrective feedback (FCF) versus Structured and Supported Debriefing (SSD)) on team dynamics in simulation-based cardiac arrest team training	A pilot randomized controlled study. Primary outcome: improvement in team dynamics scores between baseline and test simulation. Secondary outcomes: improvements in team clinical performance scores, self-assessed comprehension of and confidence in cardiac arrest management and team dynamics.	N= 95 4th year UG medical students randomly assigned to FCF or SSD; teams of 6	The SSD team dynamics score post-test was higher than at baseline [baseline: 74.5 (65.9-80.9), post-test: 85.0 (71.9-87.6), P = 0.035]. Scores for the FCF group did not improve from baseline to post-test. No differences in improvement in team dynamics or team clinical performance scores between the two groups (P = 0.328, respectively).
10	Kolbe et al. (2013) Switzerland	To describe the development of an integrated debriefing approach and demonstrate how trainees perceive this approach.	Post-test-only (debriefing quality) and a pre-post-test (psychological safety, leader inclusiveness), no-control-group design. Debriefing administered during a simulation-based combined clinical and behavioural skills training day for anaesthesia staff (doctors and nurses). Each trainee participated and observed in four scenarios and also completed a self-report debriefing quality scale.	N=61 (Four senior anaesthetists, 29 residents, 28 nurses) from a teaching hospital in Switzerland participated in 40 debriefings resulting in 235 evaluations. All attended voluntarily and participated in exchange for credits.	Utility of debriefings evaluated as highly positive, whilst pre-post comparisons revealed psychological safety and leader inclusiveness increased significantly after debriefings.

11	Lammers et al. (2012) United States	To identify causes of errors during a simulated, prehospital pediatric emergency.	Quantitative (cross sectional, observation) and qualitative research. Crews participated in simulation using own equipment and drugs. Scoring protocol used to identify errors. Debriefing conducted by trained facilitator immediately after simulated event elicited root causes of active and latent errors.	N=90 (m=67%, f=33%) Two-person crews (45 in total) made up of: EMT/Paramedic, paramedic/paramedic, paramedic/Specialist.	Simulation, followed immediately by facilitated debriefing, uncovered underlying causes of active cognitive, procedural, affective and teamwork errors, latent errors, and error-producing conditions in EMS pediatric care.
12	LeFlore & Anderson (2009)	To determine whether self-directed learning with facilitated debriefing during team-simulated clinical scenarios has better outcomes compared with instructor-modelled learning with modified debriefing.	Participants randomized to either the self-directed learning with facilitated debriefing group (Group A - 7 teams) or instructor-modelled` learning with modified debriefing group (Group B - 6 teams). Tools assessed students' pre/post knowledge (discipline-specific), satisfaction (5-point Likert scale/open ended questions), technical and team behaviours.	Convenience sample of students; nurse practitioner, registered nurse, social work, respiratory therapy. Thirteen inter-disciplinary teams participated, with one student from each discipline per team.	Group B was significantly more satisfied than group A (*P* = 0.01). Group B registered nurses and social worker students were significantly more satisfied than group A (30.0 ± 0.50 vs. 26.2 ± 3.0, *P* _ 0.03 and 28.0 ± 2.0 vs. 24.0± 3.3, *P* = 0.04, respectively). Group B had significantly better scores than group A on 8 of the 11 components of the Technical Evaluation Tool; group B intervened more quickly. Group B had significantly higher scores on 8 of 10 components of the Behavioral Assessment Tool and overall team scores.
13	Oikawa et al (2016) United States	To determine if learner self-performance	Prospective, controlled cohort intervention study.	N= 57 post graduate yr 1 medical interns	Learner SPA and TPA scores improved overall from the first

		assessment (SPA) and team-performance assessment (TPA) were different when simulation based education (SBE) was supported by self-debriefing (S-DB), compared to traditional facilitator-led debriefing (F-DB).	Primary outcome measures: SPA and TPA assessed using bespoke global rating scales with sub-domains: patient assessment, patients treatment and teamwork	randomized to 9 F-DB and 10 S-DB Teams completed 4 sequential scenarios	to the fourth scenarios (P <.05). F-DB versus S-DB cohorts did not differ in overall SPA scores.
14	Reed (2015) United States	To explore the impact on debriefing experience of three types of debrief: discussion only; discussion + blogging; discussion + journaling	Experimental design with random assignment. Primary outcome measure: Debriefing Experience Scale (DES)	N=48 UG nursing students randomly assigned to 'discussion', 'blogging' or 'journaling'.	DES score highest for discussion only, followed by journaling and then blogging. Differences reached statistical significance for only 3 of the 20 DES items.
15	Salvodelli et al. (2006)	To investigate the value of the debriefing process during simulation and to compare the educational efficacy of oral and videotape-assisted oral feedback, against no debriefing (control).	Prospective, randomized, controlled, three-arm, repeated-measures study design. After completing pre-test scenario, participants randomly assigned to control, oral or videotape-assisted oral feedback condition. Debrief focussed on non-technical skills performance followed by a post-test scenario. Trained evaluators scored participants using ANTS scoring system. Video tapes reviewed by two blinded independent assessors to rate nontechnical skills.	N=42 anaesthesia residents in postgraduate years 1, 2 and 4	Statistically significant improvement in non-technical skills for both oral and videotape-assisted oral feedback groups (*p*<0.005) but no difference between groups or improvement in control group. The addition of video review did not provide any advantage over oral feedback alone.

16	Smith-Jentsch et al. (2008)	To investigate the effects of guided team self-correction using an expert model of teamwork as the organizing framework.	Study 1: Cohort design with data collected over 2 years. Year 1, data on 15 teams collected using existing Navy method of prebriefing and debriefing. Instructors then trained using guided team self-correction method. Year 2, data collected on 10 teams, briefed and debriefed by instructors trained from Year 1. Study 2: Teams were randomly assigned to the experimental or control condition.	Study 1: n=385 male members of 25 U.S. Navy submarine attack center teams, teams ranged from 7-21 in size. Study 2: n=65 male lieutenants in the U.S. Navy, randomly assigned to 5-person teams.	Teams debriefed using expert model-driven guided team self-correction approach developed more accurate mental models of teamwork (Study 1) and demonstrated greater teamwork processes and more effective outcomes (Study 2).
17	Van Heukelom et al.(2010)	To compare two styles of managing a simulation session: post-simulation debriefing versus in-simulation debriefing.	Observational study with a retrospective pre-post survey (using 7-point Likert scale) of student confidence levels, teaching effectiveness of facilitator, effectiveness of debriefing strategy and realism of simulation. Participants randomly assigned to either post-simulation or in-simulation debriefing conditions.	N=160 students (third year medical students enrolled in the 'Clinical Procedures Rotation')	Statistically significant differences between groups. Students in the post-simulation debriefing ranked higher in measures for effective learning, better understanding actions and effectiveness of debrief.
18	Zinns et al (2017) United states	To create and assess the feasibility of a post resuscitation debriefing framework (REFLECT)	Feasibility pretest-posttest study. Outcome measure: presence of REFLECT components as measured by the PEM fellows, team members and blinded reviewers.	N= 9 Pediatric Emergency Medicine (PEM) fellows completed the REFLECT training (intervention) and led teams of 4.	Significant improvement in overall use of REFLECT reported by PEM fellows (63% to 83%, p<0.01) and team members (63% to 82%, p<0.001). Blinded reviewers found no statistical improvement (60% to 76%, p = 0.09).

Table 4 Debriefing Frameworks and Measures used in the 18 studies

Reference	*Debriefing framework*	*Outcome measure*	
		Quality of debrief	*Participant Performance*
Boet et al. (2013)	Ottawa Global Rating Scale		Team Emergency Assessment Measure
Bond et al. (2006)	Technical/knowledge (B) Cognitive (B)	Survey/ Interview (B)	
Brett-Fleegler et al. (2012)	Debrief framework to show (i) superior, (ii) average, (iii) poor debriefing (B)	DASH	
Freeth et al. (2009)	Structured (B)		Kirkpatrick framework adapted for IPE
Forneris et al. (2015)	Debriefing for Meaningful Learning (DML)	DASH	Health Sciences Reasoning Test
Geis et al. (2011)	Structured (B)	Survey (B)	Mayo High Performance Teamwork Scale
Grant et al (2014)	Video-assisted oral debriefing (B) Oral debriefing alone (B)		Behaviours (B)
Hull et al (2017)	Structured (B)	OSAD	
Kim et al (2017)	Focused & Corrective Feedback (B) Structured & Supported Debriefing		Team Dynamics Team Clinical Performance
Kolbe et al. (2013)	TeamGAINS	Survey based on DASH and OSAD	Psychological safety Leader inclusiveness
Lammers et al. (2012)	Structured (B)	Interview (B)	
LeFlore & Anderson (2009)	Facilitated debrief (B) Modified debrief (B)	Survey (B)	Knowledge Assessment (B) Technical Evaluation (B) Behavioural Assessment
Oikawa et al (2016)	Facilitator-led Debriefing (B) Self-Debriefing (B)		Self-performance assessment (B) Team-performance assessment (B)
Reed (2015)	Discussion debrief (B) Discussion + Journal (B) Discussion + Blog (B)	DES	
Salvodelli et al. (2006)	Structured (B)		ANTS
Smith-Jentsch et al. (2008)	Guided Team Self-Correction		Mental models of teamwork (B) Teamwork processes (B)
Van Heukelom et al. (2010)	In-simulation debriefing (B) Post simulation debriefing (B)	Survey (B)	Self-reported Confidence (B)
Zinns et al (2017)	REFLECT (B)	REFLECT criteria (B)	
B= bespoke			